شنبه 08 فروردین 1405 - 09:11

هوش مصنوعی همه ما را شبیه به هم می‌کند!

شواهد نوظهور نشان می‌دهد که خروجی مدل‌های زبانی بزرگ می‌تواند متن و حتی افکار کاربران انسانی را شکل دهد.

شواهد نوظهور نشان می‌دهد که خروجی مدل‌های زبانی بزرگ می‌تواند متن و حتی افکار کاربران انسانی را شکل دهد.

به گزارش ایسنا، در قلب مدل های هوش مصنوعی امروزی، مجموعه های عظیمی از داده های آموزشی قرار دارد. متن ها، ویدئوها و تصاویر ساخته شده توسط انسان ها که برای آموزش مدل ها به کار می روند تا الگوها را شناسایی کرده و محتوا تولید کنند. بدون شک انسان ها در حال آموزش سیستم های هوش مصنوعی هستند، اما آیا هوش مصنوعی نیز در حال آموزش دادن ماست؟

به نقل از نیچر، تعداد رو به رشدی از مقالات گزارش می دهند که افراد تمایل دارند الگوهای نوشتاری، روش های استدلال و حتی دیدگاه ها را از مدل های زبانی بزرگی (LLM) که استفاده می کنند، دریافت کنند. برخی پژوهشگران می گویند این تأثیر می تواند به نوعی یکنواختی در نوشتار انسانی منجر شود و هشدار می دهند که این اثر حتی ممکن است به متونی که توسط افرادی نوشته می شود که مستقیما از هوش مصنوعی استفاده نمی کنند نیز سرایت کند.

ژیوار سوراتی، دانشمند علوم رایانه در دانشگاه کالیفرنیای جنوبی در لس آنجلس و یکی از نویسندگان مقاله ای در این مورد استدلال می کند که مدل های زبانی بزرگ در حال همگن سازی گفتمان انسانی هستند و می گوید: اگر اطرافیان شما با این مدل ها تعامل داشته باشند و سبک نوشتاری، دیدگاه ها و شیوه های استدلال آن ها را بپذیرند، در نهایت این الگوها آن قدر شما را احاطه می کنند که به نظر می رسد این روش، شکل اجتماعی درست برای بیان اطلاعات است.

با این حال، برخی دیگر معتقدند ذهن انسان ممکن است همچنان در برابر این اثر هموارکننده هوش مصنوعی مقاومت کند. در یک مطالعه که در ماه نوامبر به صورت پیش چاپ در سرور arXiv منتشر شد، نویسندگان گروه هایی از نویسندگان را شناسایی کردند که «نشانه های سبکی متمایز و انسانی» خود را حفظ می کنند و احتمالاً اصالت را به سود کارایی ارائه شده توسط هوش مصنوعی ترجیح می دهند. این مطالعه هنوز داوری همتا نشده است.

بررسی عمیق تر موضوع

در یک پیش چاپ دیگر که سال گذشته در arXiv منتشر شد و هنوز داوری نشده، سوراتی و همکارانش پست های ردیت، محتوای خبری و مطالعات پیش چاپ را پیش و پس از راه اندازی چت جی پی تی در سال ۲۰۲۲ تحلیل کردند. گروه پژوهشی دریافت که متن های منتشرشده پس از عرضه این پلتفرم، از نظر سبک نوشتاری تنوع کمتری نسبت به متن های قبل دارند.

در مقاله ای جدید، نویسندگان استدلال می کنند که این پدیده بر دیدگاه ها و شیوه های استدلال افراد نیز تأثیر می گذارد. آن ها به یک پیش چاپ بدون داوری در سال ۲۰۲۳ اشاره می کنند که در آن شرکت کنندگان با مدل های زبانی ای تعامل داشتند که دیدگاه های مثبت یا منفی درباره شبکه های اجتماعی بیان می کردند. پس از این مواجهه، دیدگاه های خود شرکت کنندگان به سمت دیدگاه های تولیدشده توسط مدل ها تغییر کرد.

اولیور هاوزر، پژوهشگر اقتصاد و هوش مصنوعی در دانشگاه اکستر بریتانیا، می گوید نویسندگان نکته قابل قبولی مطرح می کنند که افراد می توانند از هوش مصنوعی سود ببرند و آن این است که به شما کمک می کند بهتر بنویسید و برای دیگران قابل فهم تر باشید. اما او هشدار می دهد: به محض اینکه این پذیرش فراگیر شود، این جمع است که بیشترین آسیب را می بیند.

ورود به عرصه سیاست

در مطالعه ای که به تازگی در مجله Science Advances منتشر شده، پژوهشگران دریافتند که دیدگاه های افراد درباره مسائل اجتماعی شروع به بازتاب دیدگاه هایی می کند که از یک ابزار هوش مصنوعی دریافت کرده اند. شرکت کنندگان از دستیارهای هوش مصنوعی برای نوشتن درباره موضوعات اجتماعی-سیاسی مانند مجازات اعدام استفاده کردند. پس از آن، نگرش های آن ها بیشتر به آنچه مدل های زبانی نوشته بودند در مقایسه با گروه کنترل که از هوش مصنوعی استفاده نکرده بودند، شبیه شد.

استرلینگ ویلیامز-سسی، یکی از نویسندگان این مطالعه و دانشمند اطلاعات در دانشگاه کرنل در ایتاکا، نیویورک، می گوید این اثر می تواند در نهایت تنوع دیدگاه های سیاسی را نیز کاهش دهد. البته میزان این تأثیر به گرایش هایی بستگی دارد که مدل های مختلف زبانی از خود نشان می دهند.

نکته مهم این است که شرکت کنندگان متوجه نشدند که تحت تأثیر چت بات ها قرار گرفته اند. حتی زمانی که به آن ها گفته شد هوش مصنوعی ممکن است دیدگاه هایشان را سوگیرانه کند، نتایج تغییری نکرد. ویلیامز-سسی می گوید: در حال حاضر نمی دانیم چگونه می توان از این موضوع جلوگیری کرد. و احتمالاً راه حل به سادگی ارائه یک هشدار ساده به کاربران نیست.

هاوزر همچنین می گوید یکنواختی تحمیل شده توسط ابزارهای هوش مصنوعی می تواند تفکر علمی را نیز محدود کند: ممکن است یک ایده دیوانه وار را از دست بدهیم که در ابتدا غیرمنطقی به نظر می رسد، اما در نهایت همان چیزی است که برای یک جهش علمی به آن نیاز داریم. مقاله ای که در ژانویه در نیچر منتشر شد نشان داد دانشمندانی که از ابزارهای هوش مصنوعی در پژوهش های خود استفاده می کنند، نسبت به کسانی که استفاده نمی کنند، تمایل دارند روی مجموعه محدودتری از حوزه ها تمرکز کنند.

سبک شخصی

با این حال، همه مطالعات نشان نمی دهند که استفاده از هوش مصنوعی به یکنواختی منجر می شود. مقاله ای در ماه نوامبر گزارش داد که نویسندگان انسانی هنگام تعامل با هوش مصنوعی به شیوه های متفاوتی تکامل پیدا می کنند. نوشتار برخی افراد واقعاً به مدل هوش مصنوعی شبیه تر می شود؛ در حالی که دیگران سبک شخصی خود را حفظ می کنند یا حتی سبکی ایجاد می کنند که بیش از پیش از هوش مصنوعی متمایز است.

مطالعاتی که نگرانی هایی درباره از بین رفتن تنوع در نوشتار انسانی مطرح کرده اند، باعث تردید یون وان، پژوهشگر همکاری انسان و هوش مصنوعی در دانشگاه هیوستون-داون تاون تگزاس، شد. او می گوید: ما فکر کردیم شاید این طور نباشد.
وان و همکارانش ۱۰ شخصیت هوش مصنوعی منحصربه فرد با پیشینه های فرهنگی و سبک های فکری متفاوت ایجاد کردند که طرح داستان هایی تولید می کردند و شرکت کنندگان انسانی بر اساس آن ها داستان های خلاقانه می نوشتند. تحلیل نویسندگان که سال گذشته به صورت پیش چاپ منتشر شد و هنوز داوری نشده، نشان داد که داستان های حاصل، سطحی از تنوع مشابه داستان هایی را داشتند که بدون کمک هوش مصنوعی نوشته شده بودند. وان می گوید به دلیل همین پاسخ های متنوع، این شخصیت های سفارشی می توانند راهی برای کاهش یکنواختی باشند.